

# SAZ Server/Service Update

## 17-May-2010

Keith Chadwick,  
Neha Sharma,  
Steve Timm

# Why a new SAZ Server?

- Current SAZ server (V2\_0\_1b) has shown itself extremely vulnerable to user generated authorization “tsunamis”:
  - Very short duration jobs
  - User issues condor\_rm on a large (>1000) glidein.
- This is fixed in the new SAZ Server (V2\_7\_0) using tomcat and a pools of execution and hibernate threads.
- We have found and fixed various other bugs in the current SAZ server and sazclient.
- We want to add support for the XACML protocol (used by Globus).
  - We will NOT transition to using XACML (yet).

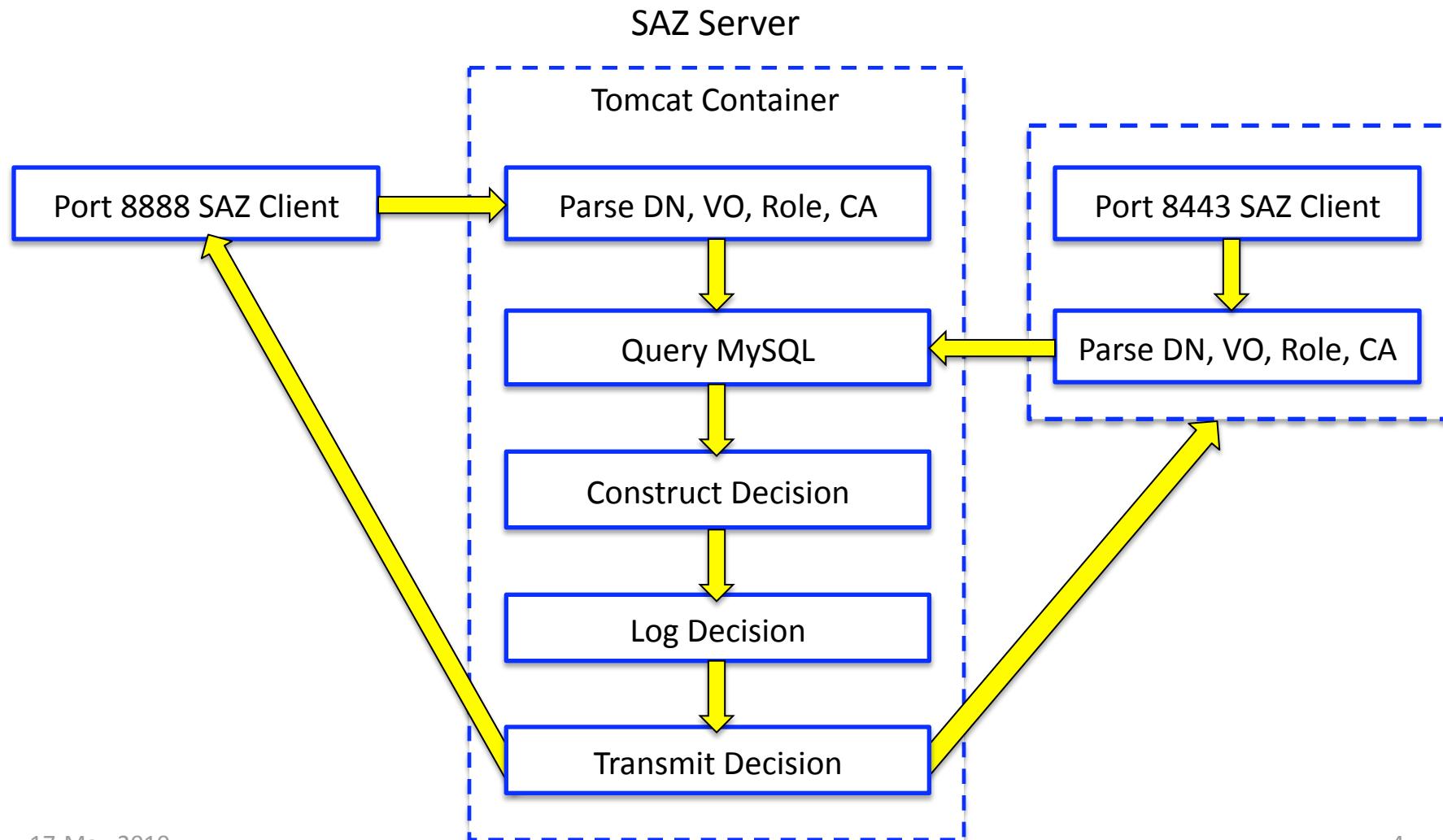
## Old (“current”) SAZ Protocol – Port 8888

- Client sends the “entire” proxy to the SAZ server via port 8888.
- Server parses out DN, VO, Role, CA.
  - In SAZ V2.0.0b, the parsing logic does not work well, and frequently the SAZ server has to invoke a shell script voms-proxy-info to parse the proxy.
  - In the new SAZ V????, the parsing logic has been completely rewritten, and it no longer has to invoke the shell script voms-proxy-info to parse the proxy.
- Server performs MySQL queries.
- Server constructs the answer and sends it to the client.

## New SAZ (XACML) Protocol – Port 8443

- Client parses out DN, VO, Role, CA and sends the information via XACML to the SAZ server via port 8443.
- Server performs MySQL queries.
- Server constructs the answer and sends it to the client.
- The new SAZ server supports both 8888 and 8443 protocols simultaneously.

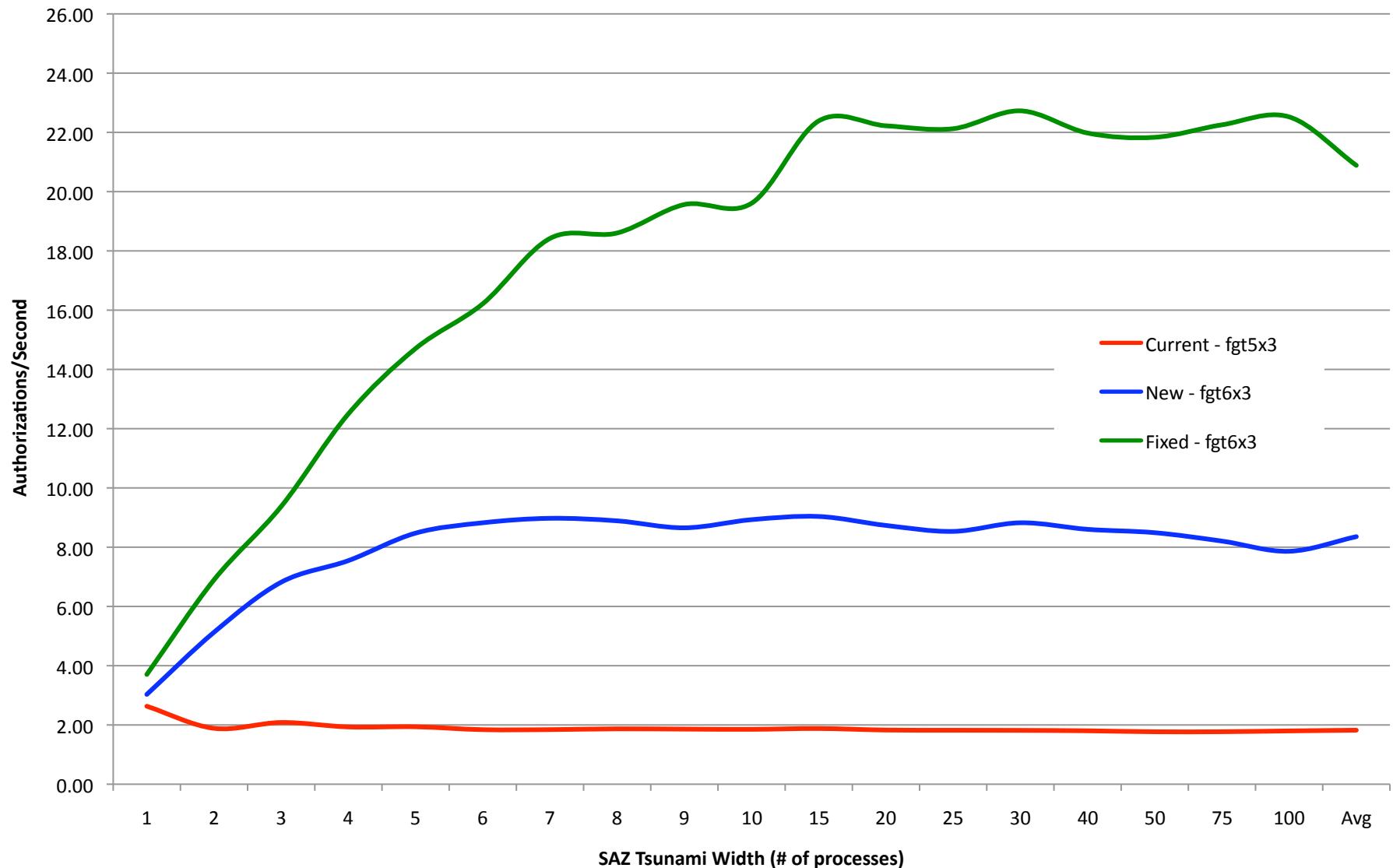
# Comparison of Old & New Protocol

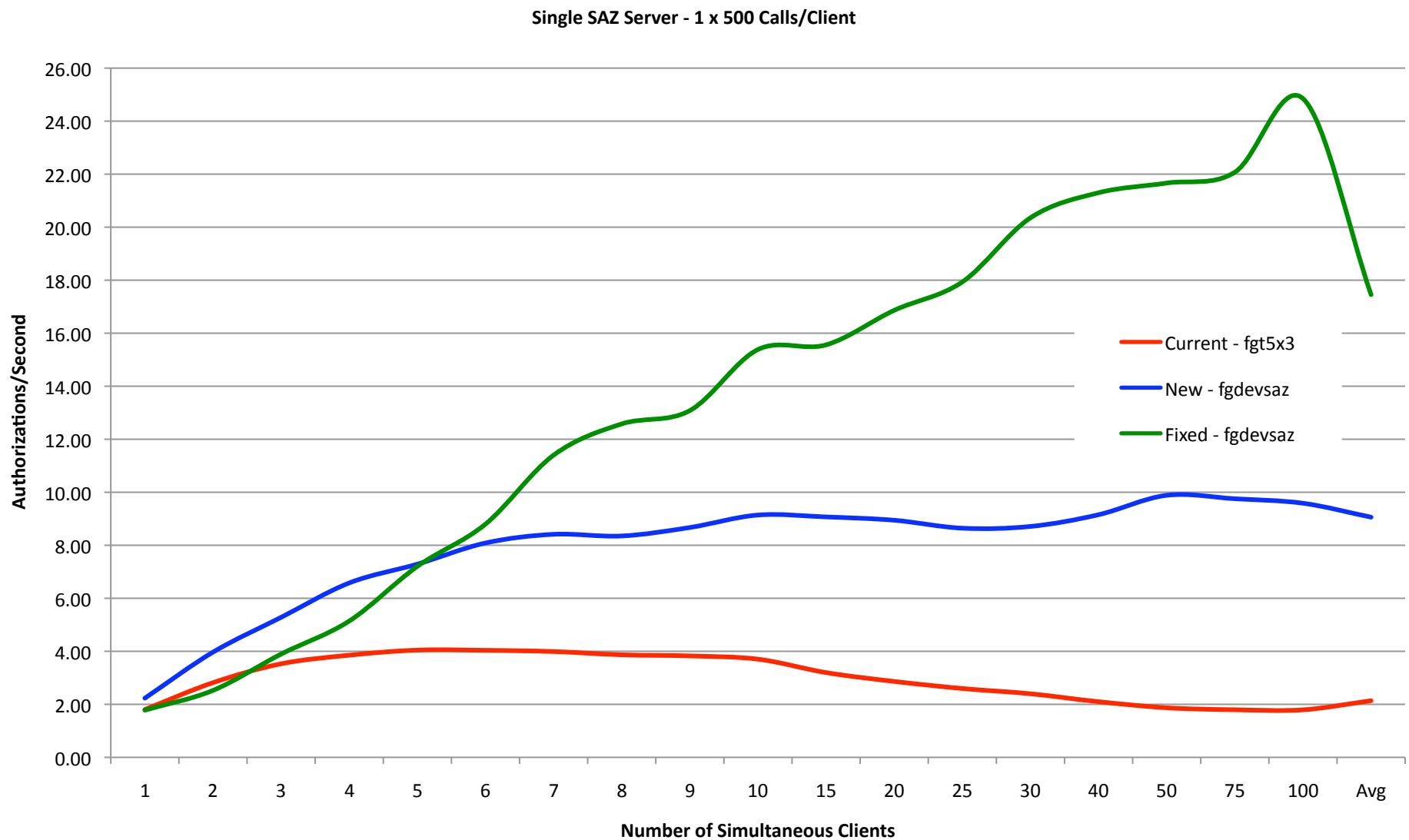


# Some Definitions

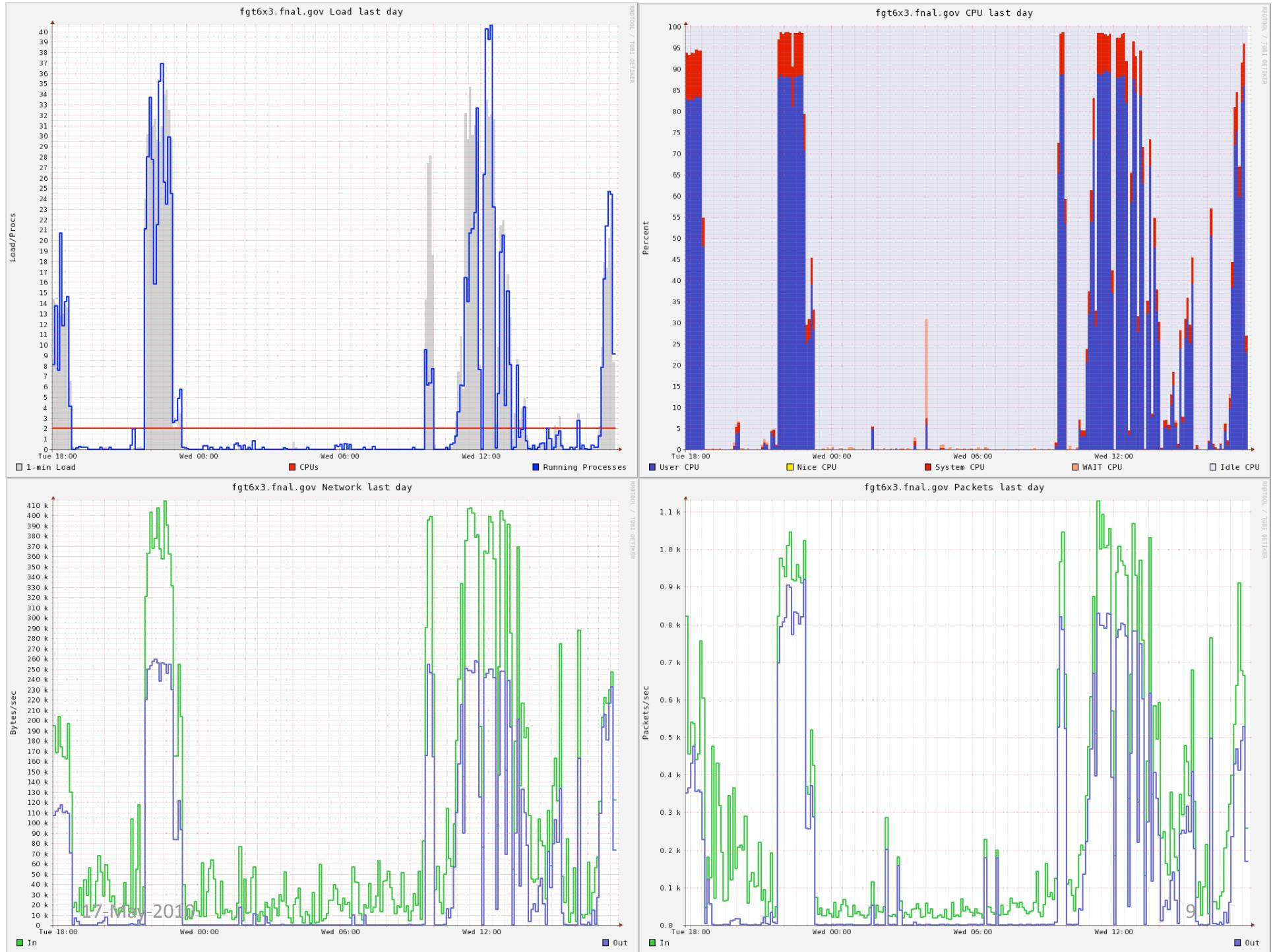
- Width = # of processes doing SAZ calls/slot or system.
- Depth = # of SAZ calls.
- Current = SAZ V2.0.0b
  - Currently deployed version of SAZ server.
- New = New SAZ Server
  - It handled small authorization tsunamis well
  - It was vulnerable to large (~1000) authorization tsunamis, (running out of file descriptors).
- Fixed = “Fixed” New SAZ Server
  - It has handled extra large (5000) authorization tsunamis without incident (ulimit 65535 to deal with the large number of open files).
  - It also has a greatly improved CRL access algorithm.
- All of the tests are run on/against SAZ servers on the fgtest systems:
  - fgtest[0-6] are FORMER production systems, 4+ years old, non-redundant.
  - Current production systems are at least 2x faster and redundant.

### Single SAZ Server Performance - Single Client System - 100 Calls/Process









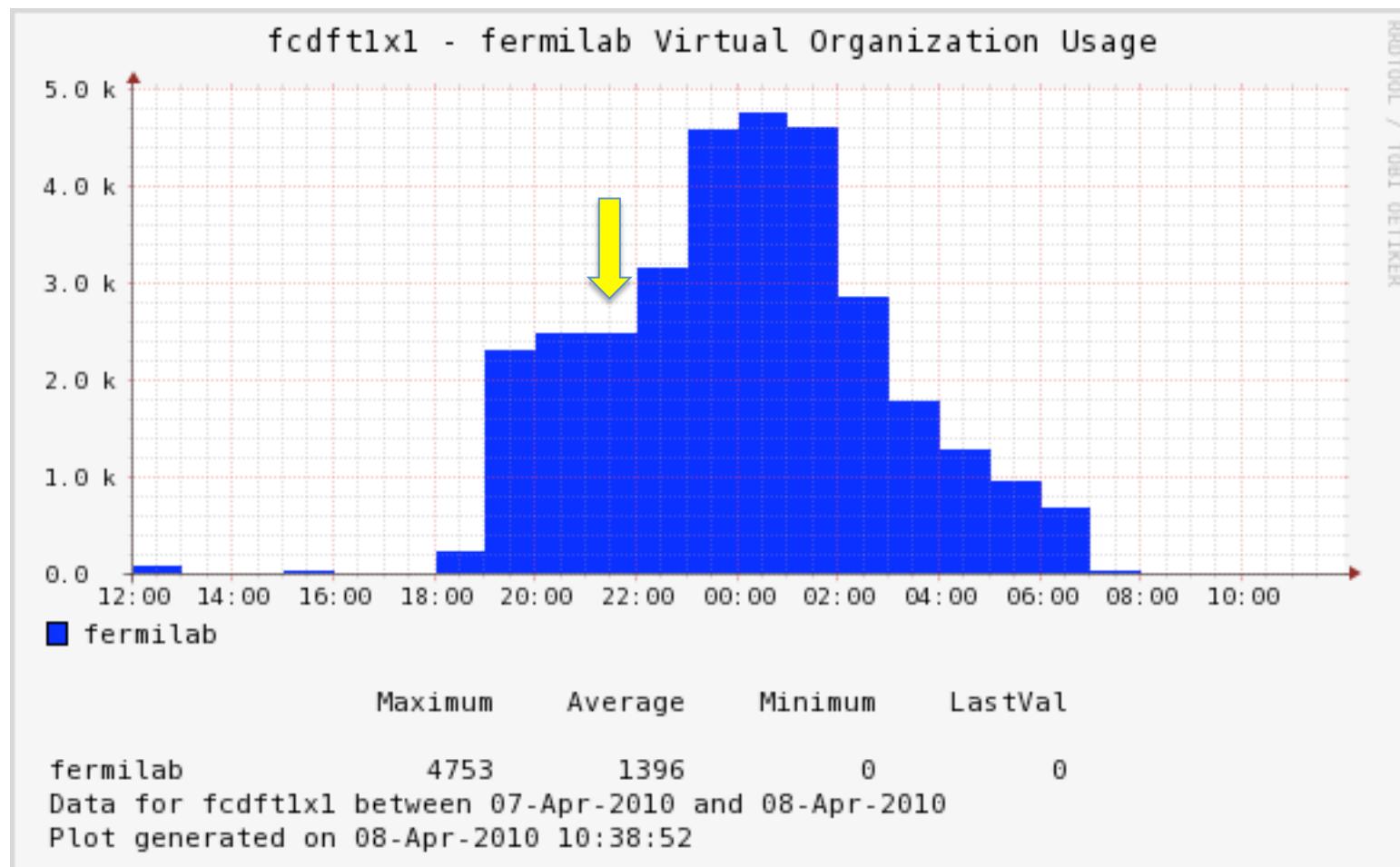
# Tsunami Testing

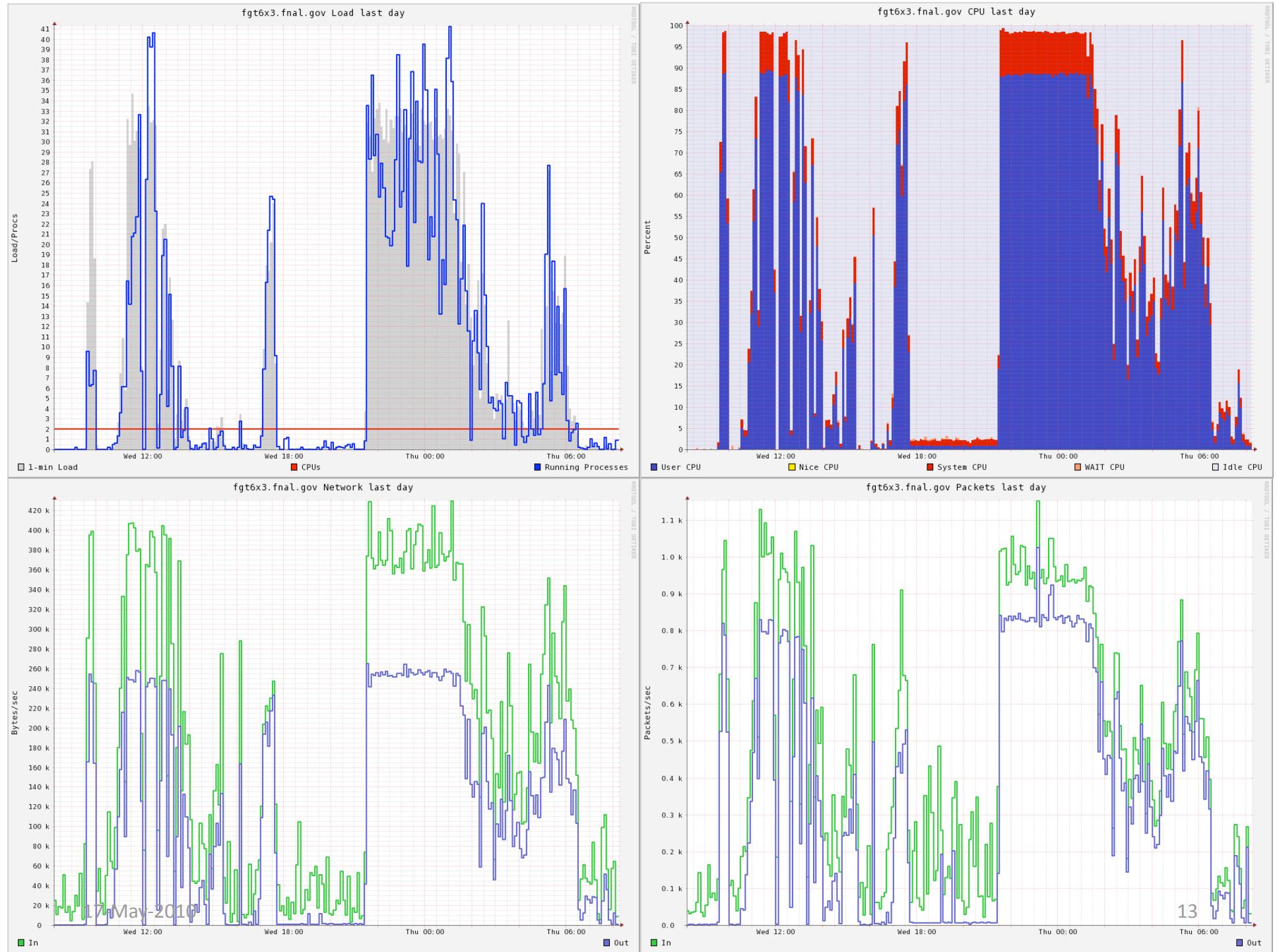
- Using fgtest systems (4+ years old).
- Submit the first set of SAZ tests:
  - jobs=50, width=1, depth=1000
- Wait until all jobs are running.
- Trigger authorizations of the first test set.
- Submit the second set of SAZ tests, either:
  - Jobs=1000, width=1, depth=50
  - Jobs=5000, width=1, depth=50
- Wait until all jobs are running.
- Trigger authorizations of the second test set.
- Measure elapsed time for first and second sets

# Results of the SAZ Tsunami Tests

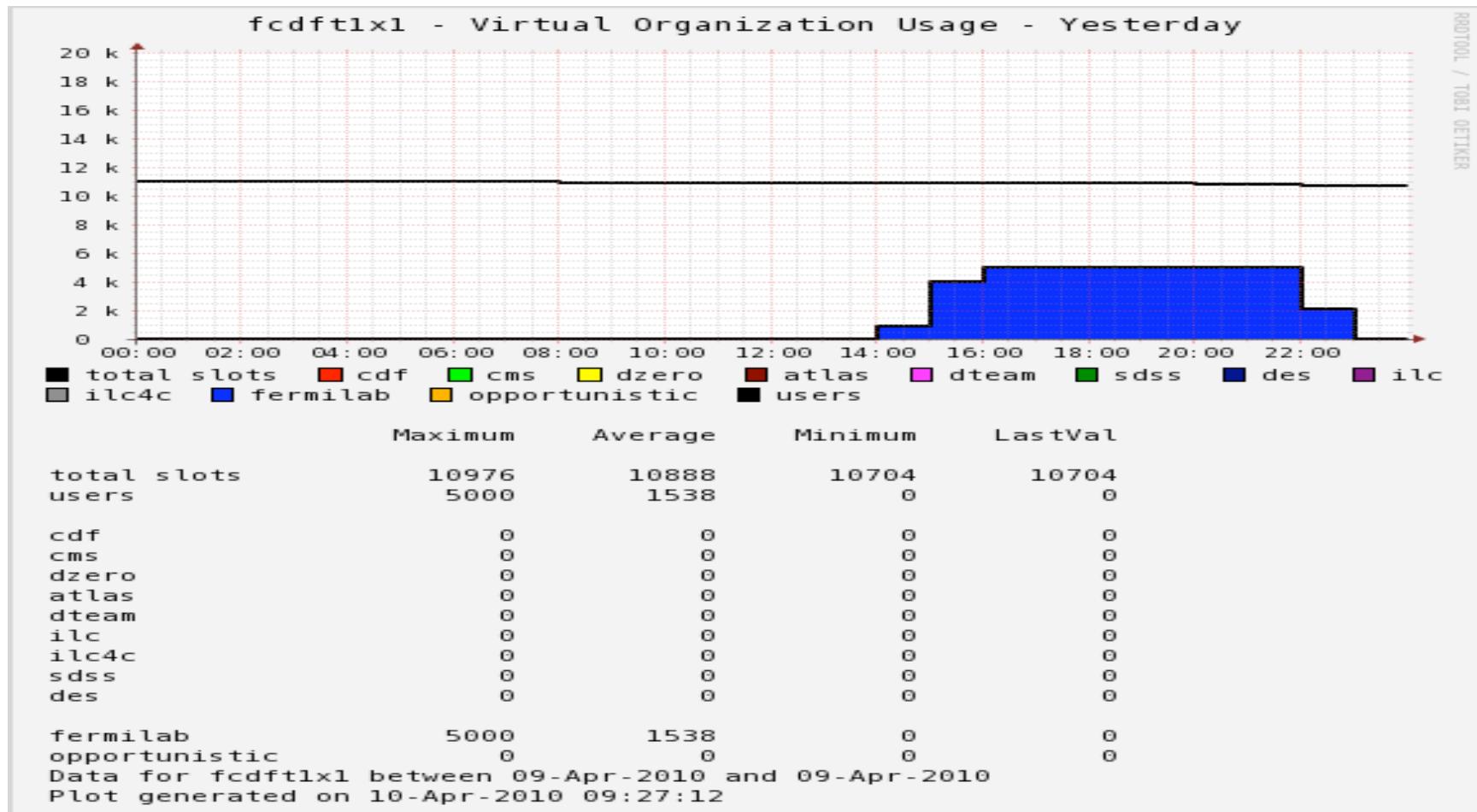
- Current SAZ (V2\_0\_1b) – fgt5x3:
  - Base=50x1x1000, Tsunami=1000x1x50
  - Immediately fail.
- New SAZ (V2\_7\_0) – fgt6x3:
  - Base=50x1x1000, Tsunami=1000x1x50
  - Ran for ~few minutes, then fail ("too many open files").
- Fixed New SAZ (V2\_7\_0) – fgt6x3:
  - Base=50x1x1000, Tsunami=1000x1x50
    - Ran without incident
    - Average of 13.68 Authorizations/second.
    - Total elapsed time was 11,205 seconds (3.11 hours).
  - Base=50x1x1000, Tsunami=5000x1x50
    - Ran without incident
    - Average >15 Authorizations/second, Peak >22 Authorizations/second.
    - Total elapsed time was ~8 hours to process base+tsunami load.

# SAZ Tsunami Profile (5000x1x50)

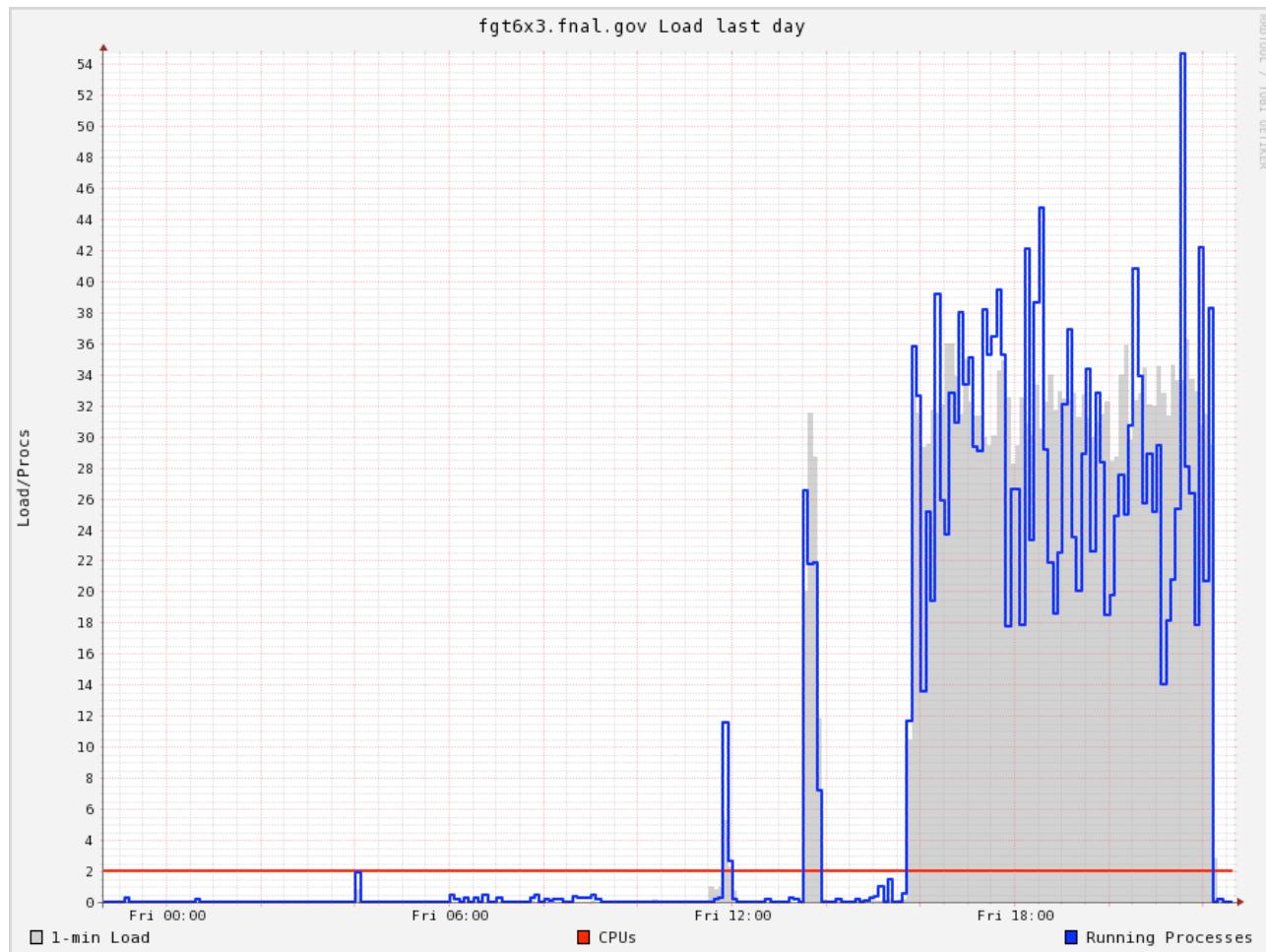




# Another Tsunami (5000x1x50)



# Load on fgdevsaz.fnal.gov

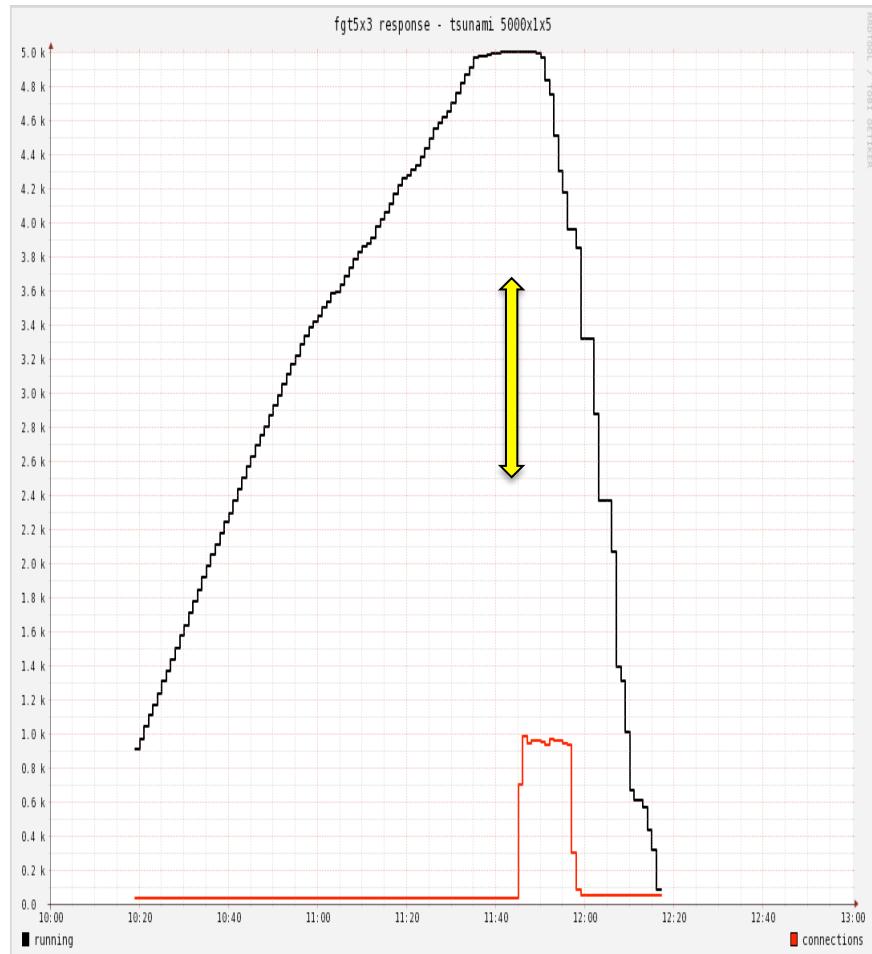


# “Real World” Scale Tsunamis

- The previous tests using  $5000 \times 1 \times 50 = 250,000$  authorizations tsunamis are well beyond the actual real world experience.
- Most authorization tsunamis have been caused by a user issuing a “condor\_rm” on a set of glide-in jobs.
- So comparison tests of fgdevsaz and fgt5x3 were run on a “real world” scale authorization tsunami –  $5000 \times 1 \times 5 = 25,000$  authorizations.

# fgt5x3 – tsunami 5000x1x5

- Black = number of condor jobs
- Red = number of saz network connections
- Trigger @ 11:45:12
- Failures start @ 11:45:19
- 25,000 Authorizations
- 14,183 Success
- 10,817 Failures
- Complete @ 11:58:28
- Elapsed time 13m 16s



# fgtdevsaz - tsunami 5000x1x5

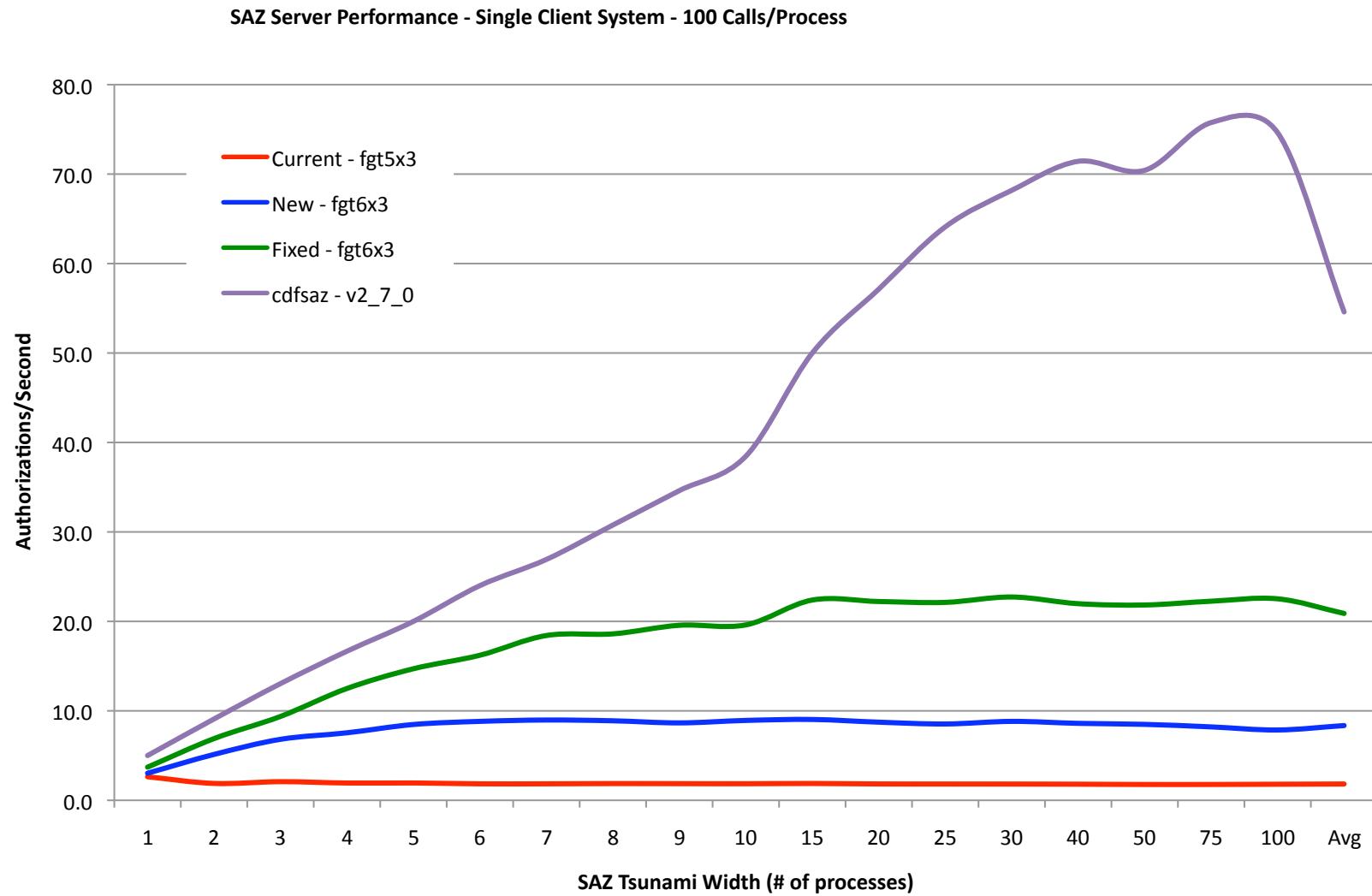
- Black = number of condor jobs
- Red = number of saz network connections
- Trigger @ 00:46:20
- 25,000 Authorizations
- 25,000 Success
- 0 Failures
- Complete @ 01:05:03
- Elapsed time = 18m 43s
- 22.26 Authorizations/sec



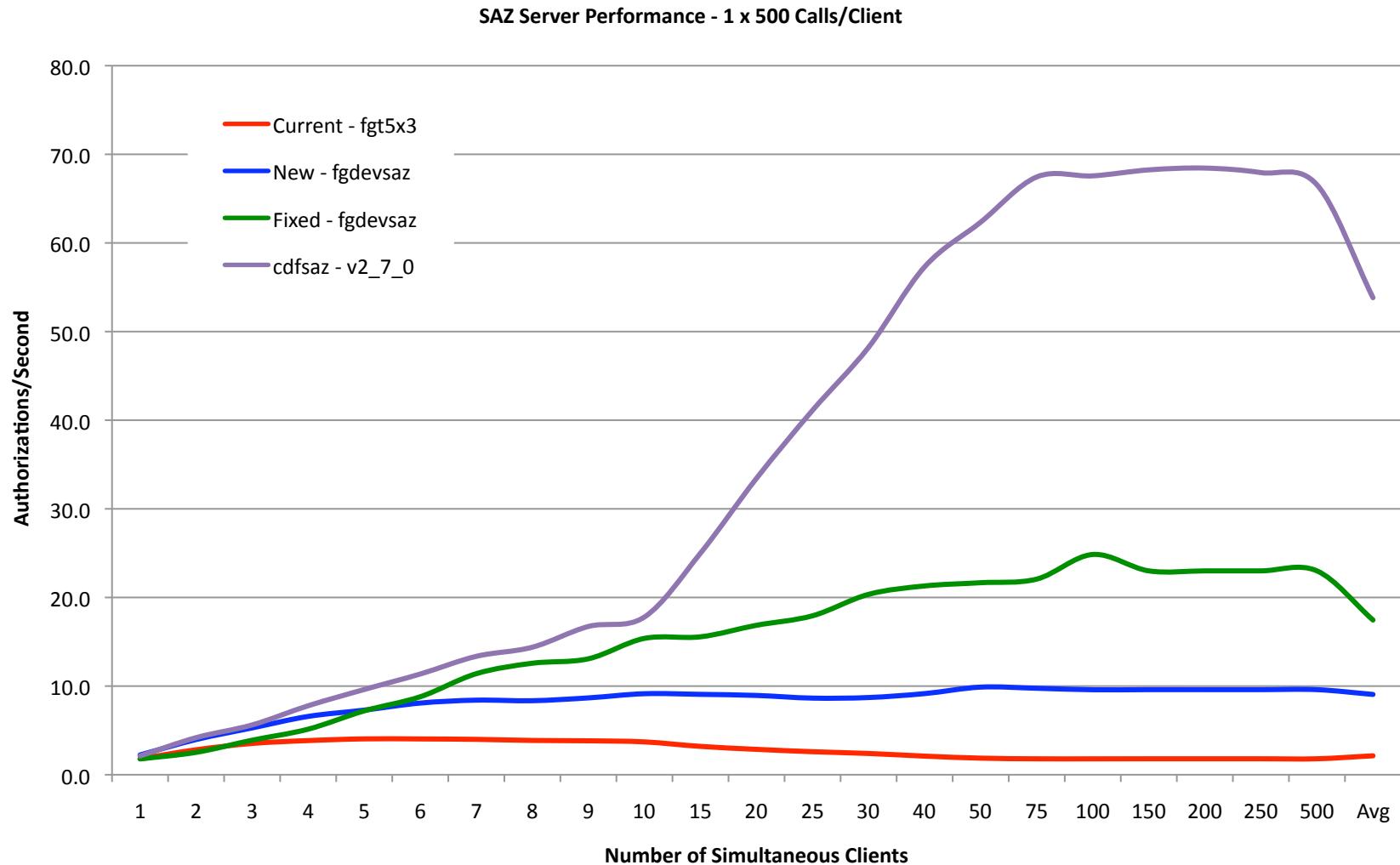
# What's Next?

- Formal Change Management Request
  - Risk Level 4 (Minor Change).
- For build & test, we propose to deploy the Fixed New SAZ service (V2\_7\_0) on the pair of dedicated CDF Sleeper pool SAZ servers. This will allow us to benchmark the Fixed New SAZ service on a deployment that substantially matches the current production SAZ service (LVS, redundant SAZ servers).
- For release, we propose to upgrade one SAZ server at a time:
  - VM Operating System “upgrade” from 32 bit to 64 bit.
  - Install of Fixed New SAZ server (V2\_7\_0).
  - Verify functionality before going to the next.

# Single client SAZ V2\_7\_0 Performance



# Multiple client SAZ V2\_7\_0 Performance



# cdfsaz:8882 – SAZ V2.7.1

## saz\_tsunami.sh cdfsaz 100 100

```
[chadwick@fgtest0 tsunami]$ ./saz_tsunami.sh cdfsaz 100 100
```

```
[./saz_tsunami.sh] - called with parameters cdfsaz 100 100 on fgtest0.fnal.gov
```

```
[./saz_tsunami.sh] - sazclient=/usr/local/vdt-1.10.1/sazclient/bin/sazclient.new
```

```
Waiting for jobs to complete...
```

```
=====
```

```
Width=100
```

```
Depth=100
```

```
Sleep=153
```

```
Calls=10000
```

```
Duration=153
```

```
Success=9960
```

```
Communication=0
```

```
=====
```

Failure rate = 40/10,000 = 0.40%

# SAZ V2.7.2

- Revised hibernate configuration
- SAZ tsunami 5000 jobs x 1 thread/job x 250 authorizations/thread

```
./saz_analyze.sh cdfsaz.5000.1.250.1273786869/
```

```
Duration - min=18646, max=19571, avg=19397.5, count=5000, total=96987420
```

```
Sleep    - min=25789, max=34153, avg=30508, count=5000, total=152539790
```

```
Success  - min=249, max=250, avg=249.999, count=5000, total=1249996
```

- Average Rate = 64.44 authorizations/second.
- 4 failures out of 1,250,000 authorizations.

# SAZ V2.7.2 Failures

```
[/grid/testhome/fnalgrid/.globus/.gass_cache/local/md5/83/41/e9/e1088254ce2aeb0d151babc322/md5/96/4c/e3/
    d229a2b12f2687ee3563796db0/data] - called with parameters cdfsaz 1 250 /grid/data/saz-tests/tsunami on fcdfcaf1002.fnal.gov
saz_tsunami.out.7.1895: PID: 29047 -- SAZ ERROR GSS.c:110 Handshake Failed... major_status != GSS_S_COMPLETE
saz_tsunami.out.7.1895: PID: 29047 -- SAZ ERROR SAZProtocol.c:64 handshake failed
saz_tsunami.out.7.1895:Thu May 13 23:58:13 CDT 2010 - sazclient=fail
saz_tsunami.out.7.1895:[/grid/testhome/fnalgrid/.globus/.gass_cache/local/md5/83/41/e9/e1088254ce2aeb0d151babc322/md5/96/4c/e3/
    d229a2b12f2687ee3563796db0/data] - results - loop=14275,call=250,pass=249,fail=1,comm=0,elapsed=19140

[/grid/testhome/fnalgrid/.globus/.gass_cache/local/md5/d9/c2/45/11a689a4b78ac9e4a3788e4c2e/md5/96/4c/e3/
    d229a2b12f2687ee3563796db0/data] - called with parameters cdfsaz 1 250 /grid/data/saz-tests/tsunami on fcdfcaf1007.fnal.gov
saz_tsunami.out.7.3169: PID: 29828 -- SAZ ERROR GSS.c:110 Handshake Failed... major_status != GSS_S_COMPLETE
saz_tsunami.out.7.3169: PID: 29828 -- SAZ ERROR SAZProtocol.c:64 handshake failed
saz_tsunami.out.7.3169:Thu May 13 22:58:18 CDT 2010 - sazclient=fail
saz_tsunami.out.7.3169:[/grid/testhome/fnalgrid/.globus/.gass_cache/local/md5/d9/c2/45/11a689a4b78ac9e4a3788e4c2e/md5/96/4c/e3/
    d229a2b12f2687ee3563796db0/data] - results - loop=8505,call=250,pass=249,fail=1,comm=0,elapsed=19414

[/grid/testhome/fnalgrid/.globus/.gass_cache/local/md5/c4/bb/1c/6653141f0f2ca340edcae2fe64/md5/96/4c/e3/
    d229a2b12f2687ee3563796db0/data] - called with parameters cdfsaz 1 250 /grid/data/saz-tests/tsunami on fcdfcaf1029.fnal.gov
saz_tsunami.out.7.3370: PID: 514 -- SAZ ERROR GSS.c:110 Handshake Failed... major_status != GSS_S_COMPLETE
saz_tsunami.out.7.3370: PID: 514 -- SAZ ERROR SAZProtocol.c:64 handshake failed
saz_tsunami.out.7.3370:Thu May 13 23:58:13 CDT 2010 - sazclient=fail
saz_tsunami.out.7.3370:[/grid/testhome/fnalgrid/.globus/.gass_cache/local/md5/c4/bb/1c/6653141f0f2ca340edcae2fe64/md5/96/4c/e3/
    d229a2b12f2687ee3563796db0/data] - results - loop=13754,call=250,pass=249,fail=1,comm=0,elapsed=19519

[/grid/testhome/fnalgrid/.globus/.gass_cache/local/md5/5f/61/35/1b92b9a1217c399b32dad9a8b6/md5/96/4c/e3/
    d229a2b12f2687ee3563796db0/data] - called with parameters cdfsaz 1 250 /grid/data/saz-tests/tsunami on fcdfcaf1049.fnal.gov
saz_tsunami.out.7.3647: PID: 17802 -- SAZ ERROR GSS.c:110 Handshake Failed... major_status != GSS_S_COMPLETE
saz_tsunami.out.7.3647: PID: 17802 -- SAZ ERROR SAZProtocol.c:64 handshake failed
saz_tsunami.out.7.3647:Thu May 13 23:58:13 CDT 2010 - sazclient=fail
saz_tsunami.out.7.3647:[/grid/testhome/fnalgrid/.globus/.gass_cache/local/md5/5f/61/35/1b92b9a1217c399b32dad9a8b6/md5/96/4c/e3/
    d229a2b12f2687ee3563796db0/data] - results - loop=12613,call=250,pass=249,fail=1,comm=0,elapsed=19235
```

# What happens at xx:58?

At 58 minutes past the hour, the hourly fetch-crl update is going against /usr/local/testgrid from fg0x6.

- We appear to have found an issue with fetch\_crl – it is supposed to move in the new CRL “safely”, but it is not working 100% when NFS is involved and 5000 clients are simultaneously using openSSL.

# SAZ V2.7.2 – Ready to Deploy

- SAZ V2.7.2 has passed all of our acceptance tests and has been demonstrated to handle authorization tsunamis that are well beyond (both in number and duration) the authorization tsunamis that have been observed “in the wild”.

# Revised Schedule of Changes

SAZ Servers	Server 1	Server 2	Status
CDF Sleeper Pool	27-Apr-2010	28-Apr-2010	Done
CDF Grid Cluster Worker Node	18-May-2010	19-May-2010	planned
GP Grid Cluster Worker Node	19-May-2010	20-May-2010	planned
D0 Grid Cluster Worker Node	25-May-2010	26-May-2010	planned
Central Gatekeeper SAZ Server	26-May-2010	27-May-2010	planned
CMS Grid Cluster Worker Node	tbd	tbd	

**Note 1:** All of the above dates are *tentative*, subject to approval by the Change Management Board and the corresponding stakeholder.

**Note 2:** FermiGrid-HA will maintain continuous service availability during these changes.